Semi-Annual Progress Report

July 1993

Mark R. Abbott

College of Oceanic and Atmospheric Sciences

Oregon State University

MODIS Team Member, Contract # NAS5-31360

Task Objectives

The objectives of the last six months were:

Review and revise the MODIS Data Products list

Continue development of local Scientific Compute Facility

Prepare white paper on ocean color measurements and distribute to the SeaWiFS science team

Complete analysis of sun-stimulated fluorescence data collected off northern California and begin for 1993 field work

Hire an information systems developer

Prepare an Algorithm Theoretical Basis Document for planned at-launch data products.

Work Accomplished

Project Data and Information System Plans

Data Products List

This activity included comments on data flows for MODIS products. The main objective was to identify data product dependencies within the MODIS product generation system. The bulk of the work on this task was completed prior to this reporting period, but final comments were provided to GSFC early this year.

The EOS Investigators Working Group (IWG) reviewed the list of MODIS data products at its March meeting. I expect that a revised list will be distributed in the next several months and further comments will be required.

Software Standards

We received a document on error handling from the MODIS SDST and provided the following comments. In general, this report was much better than the Coding Standards document that was distributed several months ago. The model implicit in this document is that processing is done on a large, central mainframe. Little attention is paid to the types of errors that arise in a networked, client/server architecture which will likely be the environment in use for EOSDIS. Issues such as resource handling, remote file systems, etc. do not appear in this draft. I think the EOSDIS project must begin to confront these sorts of issues.

The phrase "...assumes a traditional software model of a hierarchically structured, sequentially executed program with no concurrent processes." is rather limiting and not very representative of computing in the late '90's. Such a constraint limits both the types of errors and their resolution which otherwise would be much harder to manage. Again, this gets back to the whole issue of distributed applications. The Distributed Compute Environment (DCE) from OSF is a start at managing such applications. No references to external termination, deadlock, network failure, checkpointing, etc. For example, if my application requires access to a data base, which then enters a deadlock state, who is responsible for resolving the error? My application? The RDBMS?

I think the "faulty pointer manipulation" scenario is too simplistic. Other than NULL pointers, it is very hard to check the value of a pointer to see "if it is in its intended range". If it isn't, then it has the opportunity to corrupt areas of the process' memory that make any sort of controlled abort impossible (areas such as the procedure call stack come to mind).

In UNIX, trapping for certain conditions can be very expensive. Normally, in UNIX, system errors such as memory faults and arithmetic errors generate signals. The process has the option of "catching" these signals or of allowing them to perform their default action (e.g. core dump, ignore, etc.). The problem is that on some architectures, attempting to catch arithmetic errors has the unwanted effect of turning off some pipelining. The reason is that UNIX exceptions are "precise", meaning that they demand to know exactly where the exception occurred an provide a method for continuation. Unfortunately, with super-pipelined, super-scalar, out-of-order and speculative execution CPUs, the notion of "where" has gotten very fuzzy and the UNIX model of exceptions can't be supported any more. So, some care must be used when determining exactly how one is going to check for errors. Just because POSIX says you can, doesn't mean you can at 500 MFLOPS. There are numerous examples of such difficulties in the UNIX market, where high performance workstations suddenly overstress the capabilities of the operating system for error handling.

Most of the scenarios for error recovery aren't very realistic. I also think few developers will implement them. Code that aborts correctly is rare enough. The example of delay loops that sit and spin, waiting for some condition to change is pretty silly. Even MS-DOS's "Abort, Retry, Ignore" paradigm is more advanced.

This still feels like "bottom-up" design. If this were engineering instead of computing, we'd have MIL-STD designed brick with a 1000 page document detailing its dimensions, color, etc. long before we had any idea of whether we were building a house, a bridge, or a glider. The point is that while some bottom-up design is necessary, we are spending way too much time worrying the simple details and not enough time thinking about the more difficult, system design issues. This is the fundamental problem with the EOSDIS activities; focus teams are getting into the nitty-gritty way too early in the game. I don't even know what system I am supposed to be designing towards, yet I am supposed to worry about the arcane methods of error trapping. On the good side, the "single point of exit" is good design feature. There are other examples, but in general this document has a long way to go.

Local Scientific Compute Facilities

The new Environmental Computer Center at OSU was completed in late March, and most of our existing computer hardware was moved into the new facility. The Thinking Machines CM-5 was delivered on 7 April. After initial testing and debugging, the CM-5was brought on-line. Our SCF, which was funded primarily by my EOS Interdisciplinary grant, now includes two massively parallel machines (a CM-200 and the CM-5) as well as a cluster of IBM RS/6000 workstations. Although this system was designed to support ocean modeling, there is compute power sufficient to support MODIS activities as well.

As part of our joint research agreement with IBM, we expect to add a large data store device, likely one based on D2 tape technology. This system will support up to 6 TB of data and will include a hierarchical file management system. We will test the necessary infrastructure to integrate this system into our compute environment.

In terms of software, we have installed AVS on our family of UNIX workstations and are now working on integrating it with AVS on the CM-5. This system will form the basis of our visualization system. We expect to test a linked version of IDL and AVS later this summer. We developed a widget-based analysis system for our drifting buoys using IDL. This system is modular, and it should be relatively easy to extend as new functions are required. This version runs across all IDL-supported UNIX workstations as well as on Apple Macintoshes.

Otis Brown and I proposed to the Naval Research Laboratory to test advanced networking methods to develop an integrated algorithm processing and validation system for SeaWiFS. The system would be based on the Asynchronous Transfer Method (ATM) protocols that are emerging as the new standard for both wide area and local area networks. This scalable protocol is capable of handling Gbps transfer rates as well as a variety of data types, including data, audio, and video. We proposed to link OSU and the University of Miami to research facilities in the Washington D.C. area (including NASA/GSFC) using high-speed ATM lines.

Our plan was to host parts of the SeaWiFS processing system on the CM-5 at OSU and manipulate the processing system using machines at Miami. For example, a researcher at Miami could modify various parameters related to atmospheric correction and see the results in near real-time as they are produced on the CM-5. Eventually, we would extend this network to include both archive and data storage capabilities at Miami, NASA/GSFC, and OSU. As ATM devices can be incorporated into a variety of sensors, our plan was to equip drifting buoys in the ocean with ATM links so that they would appear as one more device on the network.

Presently, NRL is funding the ATM switches at Miami and OSU for local networking. We are working on acquisition of the high-speed lines necessary to link the two institutions. Some progress is being made on this front, but we do not expect a WAN implementation until sometime in 1994,

With the support of Hewlett-Packard and Ellery Systems, Berrien Moore (University of New Hampshire) and I are developing a prototype data system to test software from Ellery Systems that is based on DCE (Distributed Compute Environment). This software simplifies the programming required to develop distributed applications. The prototype system will involve subsets of Moore's Landsat data and my AVHRR/CZCS data. We will develop an application that will support distributed researchers to search through and browse our data holdings through a common interface. Simple analysis tools will also be available.

Software Development and Data Plan

Bob Evans and I are still tasked with producing a Software Development and Data Plan for the MODIS Oceans Team. Thus far, Evans has not produced a draft.

I was made a member of the EOSDIS Data Processing Focus Team two days before the first meeting so I was unable to attend. My reaction to the first report is that it has put the cart before the horse. No one seems to be able to describe the data flows at even the most basic level, yet the project is trying to identify word processors as part of its data processing strategy.

Ocean Color White Paper

The Ocean Color White Paper was distributed to the U.S. investigators on the SeaWiFS Science Team in April after an oral presentation to the Science Team at its January meeting. So far, I have received comments from Frank Hoge, Curt Davis, Ian Barton, John Parslow, Frank Müller-Karger, and Ghassem Asrar. I am disappointed with the breadth of response, as I feel this document could set the direction for ocean color research for the next decade.

The comments thus far identified management issues that need to be clarified (e.g., the status of an NRA for EOS-Color) as well as some technical issues related to sea surface temperature. One noticeable omission is the whole area of software development. Presently, NASA is supporting several related efforts for SeaWiFS, MODIS, and Pathfinder, yet one of the primary concerns at the SeaWiFS Science Team meeting was the nature of the software that would be available to the community. One of the reviewers has agreed to provide some text on this issue, and it will be incorporated in the final version.

As part of this report, it became clear that there needed to be a meeting between the NASA program and project management and some science representatives to discuss technical and management issues related to MODIS and SeaWiFS. I sent a letter to Ghassem Asrar after receiving the approval of the MODIS Oceans Team. The letter requests that a meeting between the principals take place as soon as possible. So far, I have received no response from NASA.

Data Analysis and Interpretation

The previous semi-annual report contained the major findings from analysis of bio-optical data collected from a Lagrangian drifter. A manuscript is now circulating for comments from my co-authors.

As described in the last semi-annual report, a set of Lagrangian drifters has been deployed in the California Current as part of a study funded by the Office of Naval Research. The drifters include a spectroradiometer built by Satlantic (Canada) which will measure water-leaving radiance at the SeaWiFS wavelengths as well as at 683 nm. The drifters are the standard World Ocean Circulation Experiment (WOCE) design, and the spectroradiometer is installed in the surface float. The data are relayed back via Service Argos.

The first drifter was deployed on 5 May 1993 off northern California, and it continues to operate. It has traveled several hundred kilometers and is now south of Monterey. Three more were deployed along 39.5°N in early June, and these continue to work quite well. I have included figures that show the drifter tracks and some representative bio-optical data.

Of particular interest is the sun-stimulated fluorescence data. In general, the patterns are in line with our expectations; as the water becomes bluer, the fluorescence decreases. The final figure shows the relationship between sun-stimulated fluorescence (normalized by incoming photosynthetically available radiation) and a green/blue ratio (555nm/443nm). There is a rough linear relationship, but there is clearly considerable scatter. However, as we partition the data to various water types, we expect the scatter to decrease.

Six drifters will be deployed in early July to investigate decorrelation scales of bio-optical properties. I plan to continue deployments through spring 1994. Unfortunately, with the delay of SeaWiFS it is unlikely that there will be any overlap between these deployments and SeaWiFS.

Analysis plans are focusing on estimating Lagrangian decorrelation scales for upper ocean bio-optical properties and for comparing sun-stimulated fluorescence with primary productivity measurements.

Hiring of Information Systems Developer

I have hired Roen Hogg to fill this position, and he will begin on 2 August 1993. His background is in object-oriented design, and he worked for the Advanced Technology Group of American Express. He has considerable experience in data base and knowledge management and in running software teams. I am confident that he will become a pivotal person of my team.

Algorithm Theoretical Basis Document

The MODIS Team Leader has requested that each team member prepare an ATBD describing the scientific heritage of each data product. This document should include a description of the algorithm as well as an estimate of the errors as well. I am in the process of developing the ATBD for my at-launch product, chlorophyll fluorescence line height. I am collaborating with Wayne Esaias on the ATBD for primary productivity.

Anticipated Future Actions

I will continue to review various EOSDIS documents as they appear. I hope that the EOSDIS Core System contractor will begin to contact Team members; such contacts are essential if the proper system is to be built. At this point, we have had no formal contacts. I am very uneasy about the pace of design as many requirements are being "locked in" with little input from the non-NASA science community. In addition, many standards, such as data formats, are being decided where there is no urgency for such decisions.

I will continue to analyze the ocean drifter data, especially the data on sun-stimulated fluorescence. I plan to have a manuscript submitted for publication by the end of the year.

With Mr. Hogg, I will develop a prototype system to link satellite imagery, ocean drifter data, and visualization/analysis tools in an end-to-end data management system. Mr. Hogg will investigate various data base systems, including products from the object-oriented data base companies. The intent is to develop a system that keeps track of data through every step of the collection/processing/analysis procedure. Although audit trails are fairly common for processing, this link is usually lost in the analysis stage. For example, one should be able to determine the heritage of the data displayed in an x-y plot. We will use the drifter study described earlier as the source data set as it includes satellite imagery, meteorological data, and bio-optical measurements.

In collaboration with Bob Evans and Art Mariano (Univ. Miami), we are investigating installation of Miami's optimal interpolation routines on our CM-5. These algorithms use information concerning the underlying time/space variability of pigment to fill in gaps in CZCS imagery caused by clouds, orbital characteristics, and other processes. Although we had intended to port SeaWiFS algorithms to the CM-5, this issue of interpolation was determined to be of higher priority.

Problems and Solutions

I am concerned that the EOSDIS design is proceeding without adequate scientific input from outside NASA. Thus far, no one from the ECS contractor has contacted me, let alone describe what ECS will be. Many issues are being decided without any sense of their priority or importance to the overall program. I worry that we are focusing on the insignificant details and losing sight of the larger, more complex issues. I think a solution must involve more communication between the project and the research community. These issues are clearly spelled out in the latest MODIS Science Team minutes. In effect, concrete is being poured, yet we do not have any blueprints.

I need to hire a postdoctoral researcher to assist me with analysis of the fluorescence observations. If MODIS funding is close to that which was planned, then I should be able to hire someone in FY1994. The shortage of scientific personnel is becoming acute. I have proposed some post-launch products (near-surface chlorophyll and productivity via fluorescence) that need considerable field and lab work before they will be ready to deliver to EOSDIS. Also, the basic fluorescence measurements need further study from a phytoplankton physiology point of view. I have made contact with one potential postdoctoral researcher who has considerable experience in this area.

Finally, I am concerned about the assembly of a coherent ocean color data set from the numerous sensors that will be launched in the next ten years. This is a unique opportunity, but there are significant challenges in terms of data access, consistent processing, cross-calibration, etc. I hope that NASA and the ocean color community will soon come to closure on these important issues.